Uncertainty Management in a Pyramid Vision System

Ze-Nian Li

ABSTRACT

This paper presents an efficient adaptation and application of the Dempster-Shafer (D-S) theory of evidence, one that can be used effectively in a massively parallel hierarchical system for real-world image analysis. It describes the techniques used and shows some extended examples that use a multilevel set of processes. The knowledge representation technique employs uncertain and incomplete world knowledge. It is modular, flexible, and easy to update. The reasoning mechanism extends the applications of the belief functions and Dempster's combination rule in a relatively efficient and powerful manner. The performance of the proposed D-S evidential reasoning mechanism is compared with other methods based on the Bayesian formalism and a simple weight combination method.

KEYWORDS: ***evidential reasoning, image analysis, pyramid, representation, uncertainty***

INTRODUCTION

We have been developing a computer vision system that is implemented in a pyramid-like structure that would eventually consist of many thousands of computers. The system has been effective for analyzing images containing complex objects like houses and neurons (Li and Uhr [1, 2]). This paper describes a new evidential reasoning mechanism based on the Dempster–Shafer (D-S) theory of evidence that we have recently developed for and incorporated into the pyramid vision system.

Basically, the evidential reasoning is attempting two tasks: evidence accumulation and hypothesis verification. Uncertainty and incompleteness are inevitable in both phases. Because of the fact that thousands of processors will be employed in the pyramid multicomputer, these processors are necessarily very simple and without too much local memory space. The knowledge representation and the

Address correspondence to Professor Ze-Nian Li, School of Computing Science, Simon Fraser University, Burnaby, BC, Canada V5A 1S6.

International Journal of Approximate Reasoning 1989; 3:59–85

655 Avenue of the Americas, New York, NY 10010 0888-613X/89/$3.50

reasoning mechanisms must be simple and modular, so that they can be manipulated by the massively parallel processors in the pyramid.

In the past few years, the D-S theory has been adapted in several application domains (Garvey et al. [3], Lowrance [4], Gordon and Shortliffe [5], Lu and Stephanou [6], Wesley and co-workers [7, 8], Zhang and Chen [9]).

In his dissertation [4], Lowrance defined a *dependency-graph model.* The assumption was that the domains of interest can be modeled in a propositional framework. The frame of discernment was represented by dependency graphs, where the nodes represented propositions and the arcs represented relationships, such as *and,* or, $\rightarrow$, *not,* between the propositions. Based on the Dempster-Shafer theory of evidence, Lowrance claimed that a formal system capable of pooling and extending evidential information could be built, while maintaining internal consistency.

The early version of the dependency-graph model approach appeared in 1981, when Garvey et al. [3] presented a program that was capable of identifying electromagnetic signal emitters from the measurement of characteristics of the signals emitted by these devices. The program directly used Dempster's combination rule for its reasoning process. Wesley and co-workers [7, 8] elaborated the dependency-graph model approach, especially in the domain of evidential knowledge-based computer vision.

The dependency-graph model approach has been successfully demonstrated in several domains with small sets of objects. It has the merit of being formal and sound. However, the success of the dependency-graph models relies on the construction of a fairly complete graph. It would be difficult to extend to more general and complex domains.

Lu and Stephanou [6] presented a different framework for the processing of uncertain knowledge using the D-S theory. Evidence (e) and hypothesis (h) are viewed to be in two separate spaces—input space I and output space O. A set of mappings is defined between these two spaces. A mapping is expressed as $e \rightarrow h, v$, where $e \subset I$, $h \subset O$, and $0 \leq v \leq 1$. The variable v represents the degree of association in the expert's opinion, and $1 - v$ is the degree to which the expert chooses to be noncommittal, or the degree of ignorance. This mapping is used to verify hypotheses in the evidential reasoning. First, belief functions from independent observations are combined using Dempster's rule. A combined belief function in the input space is thus generated. Second, the belief in $e \subset I$ is multiplied by the number v to get a new belief in $h \subset O$. As a result, new belief functions in the output space are generated. Third, the new beliefs are "normalized" using the Dempster rule. The concept of mapping between evidence and hypothesis spaces is very interesting. It presents a mechanism to accommodate the uncertain and incomplete nature for both the evidence and the system's knowledge. However, the simple multiplication by the number v used in the second step seems to be a weak method for accomplishing this mapping.

Recently, Zhang and Chen [9] applied the D-S theory in understanding

multispectral images of remote-sensing systems. Comparisons were made among their D-S contextual classification method, the Bayesian classification, and other stochastic relaxation methods.

This paper will describe a mechanism for uncertainty management in a pyramid vision system. The following section presents a set-theoretical evidential reasoning approach. The method for evidence accumulation is illustrated. A modular knowledge representation technique and a new mechanism for hypothesis verification are proposed. In the third section, a pyramidal object recognition approach is described and examples for uncertainty management are presented. The fourth section gives the comparative results between our D-S evidential reasoning mechanism and two other methods based on the Bayesian formalism, and the simple weight combination. The fifth section concludes the paper.

A SET-THEORETICAL EVIDENTIAL REASONING MECHANISM

This section introduces a set-theoretical evidential reasoning approach based on the Dempster–Shafer mathematical theory of evidence. The description of the Dempster–Shafer theory can be found in publications by Dempster [10] and Shafer [11].

Here, the notations in Ref. 11 are adopted.

Let Θ be a set of propositions about the mutually exclusive and exhaustive possibilities in a domain. Θ is called the *frame of discernment,* and 2^Θ is the set of all possible subsets of Θ.

A function m: $2^\Theta \rightarrow [0, 1]$ is called a *basic probability assignment* (*bpa*) if it satisfies $m(\varnothing) = 0$ and $\Sigma_{A \subseteq \Theta} m(A) = 1$.

If m is a bpa, then a *belief function* is defined by

$$\text{Bel}(A) = \sum_{B \subseteq A} m(B), \qquad \text{for } A \subseteq \Theta$$

Evidence Accumulation

The accumulation of evidence is conducted in two steps, for single-feature assessment and for multifeature combination. Features extracted from the real-world image data are usually inaccurate and incomplete. The uncertainty here represents the uncertainty in the evidence.

SINGLE-FEATURE ASSESSMENT At this step, probability values are assigned to the extracted features to represent their uncertainty. The following factors are taken into account:

1. The quality of the input data and the weights of the extracted features.
2. The "goodness" of feature values. A feature value is assessed by

comparison with the typical feature value of a hypothesized object. Questions like "How good is the shape of this *long* region, if it is compared to the shape of a typical window shutter?" are asked in the assessment.

These factors are combined into a single probability value. As a result, each feature f gets a probability mass $0 \leq m(f) \leq 1$. In most cases, $m(f)$ is less than 1, and the remainder of the unit mass $1 - m(f)$ is assigned to Θ to represent the uncertainty. Therefore, for single-feature assessment, *simple belief functions* are used, where $\text{Bel}(f) = m(f)$.

MULTIFEATURE COMBINATION If each extracted feature is viewed as a piece of evidence, then the independent observations on multifeatures can be combined to serve as accumulated evidence. Belief functions of independently extracted features are combined using Dempster's combination rule.

Suppose m_1 is the bpa for a belief function Bel_1 over a frame Θ, with focal elements A_i $(i = 1, \ldots, k)$; and similarly m_2 is the bpa for a belief function Bel_2 over the same frame Θ, with focal elements B_j $(j = 1, \ldots, l)$.

If $K = \Sigma_{A_i \cap B_j = \varnothing} m_1(A_i) m_2(B_j) < 1$, then the function m: $2^\Theta \rightarrow [0, 1]$ is a bpa, defined by

$$m(\varnothing) = 0, \qquad \text{and} \qquad m(C) = (1 - K)^{-1} \sum_{A_i \cap B_j = C} m_1(A_i) m_2(B_j)$$

The belief function m that combines Bel_1 and Bel_2 is called the *orthogonal sum* of Bel_1 and Bel_2, denoted $\text{Bel}_1 \oplus \text{Bel}_2$. Since Dempster's combination rule is associative and commutative, features can be combined in any order.

A Modular Knowledge Representation Technique

In the field of computer vision, one of the popular knowledge representation schemes used in evidential reasoning is the dependency-graph model [4, 8]. Due to the complexity of the graph, the world model would take potentially very large amounts of space to store and time to process, and it would be difficult to update. As Lowrance stated [4], "as these graphs become larger and more complex, it becomes increasingly difficult to guarantee their logical consistency."

The essential point is that a knowledge representation scheme should be able to utilize incomplete knowledge. There should be a way to emphasize the important subset of the world knowledge. Truth should be expressed relative to portions of the environment, responsive to the context. Although the initial knowledge will probably generate poor results, it should be flexible and easy to improve. These all suggest that a simple and modular knowledge representation technique must be developed, especially in a massively parallel processing environment.

Our new knowledge representation technique is characterized by associating probability values with the *expected* feature components of the hypothesized objects. Probability values are used to answer questions like "If it has a typical elongated shape for a window shutter, how much should this feature of shape contribute to the belief for a shutter?" If the knowledge about window shutters is that "Window shutters are usually elongated rectangular regions, with a low 'edgeness' measure, next to windows," then the following probability mass assignments $m_s(shutter)$ may be given to represent this knowledge:

$$m(long) = 0.25, \qquad m(low) = 0.15, \qquad m(long \ \& \ low) = 0.15,$$
$$m(next\text{-}to) = 0.25, \qquad m(\Theta) = 0.2$$

To discriminate different objects in a vision system, it is often important to consider beliefs about multiple features, especially those important features of the hypothesized object. The importance of the feature is represented by the amount of the mass assigned it. The amount of $m(\Theta)$ represents the uncertainty and incompleteness of this knowledge. The numbers initially assigned may be subjective and ineffective. However, they can always be adjusted later on.

If the current task is to discriminate shutters from a chimney, and the system's knowledge about chimneys is "Chimneys are usually elongated rectangular regions, *not* next-to windows," and the system has no knowledge about the importance of the texture, then a mass assignment $m_s(chimney)$ may be given at the same time:

$$m(long) = 0.25, \qquad m(\overline{next\text{-}to}) = 0.35, \qquad m(\Theta) = 0.4$$

In this way, the knowledge representation technique allows the invocation of the subsets of the world knowledge that are most closely related to the current context of the objects being recognized and have the most discriminative power for the current hypotheses.

Hypothesis Verification

The mechanism of hypothesis verification can be viewed as a set of mappings between two spaces. The first is the evidence space E, and the second the object space O. A mapping

$$e \rightarrow o, \ \mathrm{Bel}(o)$$

defines a belief function $\mathrm{Bel}(o)$ over the object space O, where $e \subset E$ is the set of accumulated evidence, $o \subset O$ is the hypothetical object, and $0 \leq \mathrm{Bel}(o) \leq 1$ is the belief committed to o. While conducting this mapping, the knowledge sources are consulted.

As described in the previous two subsections, the accumulated evidence and the uncertainty on the evidence can be represented by a belief function whose

mass distribution is m_e, whereas the system's knowledge of the hypothetical object and the uncertainty on this knowledge can be represented by another belief function whose mass distribution is m_s. Here a third belief function Bel(o) is introduced to represent the result of the hypothesis verification.

THE bpa OF Bel(o)

$$m(o) = \sum_{\substack{A_i \subseteq B_j \\ B_j \neq \Theta}} m_e(A_i) m_s(B_j), \qquad m(\Theta) = 1 - m(o)$$

This belief function is graphically depicted in Figure 1. The areas whose masses are committed to the hypothetical object o are marked with asterisks. The value of Bel(o) is the sum of these areas. The use of the set inclusion operator in the definition of Bel(o) performs the consistency check between the system's knowledge and the evidence accumulated so far. The creation of this new belief function has extended the notions of the belief functions and the Dempster combination rule. The belief function thus generated is a simple belief function.

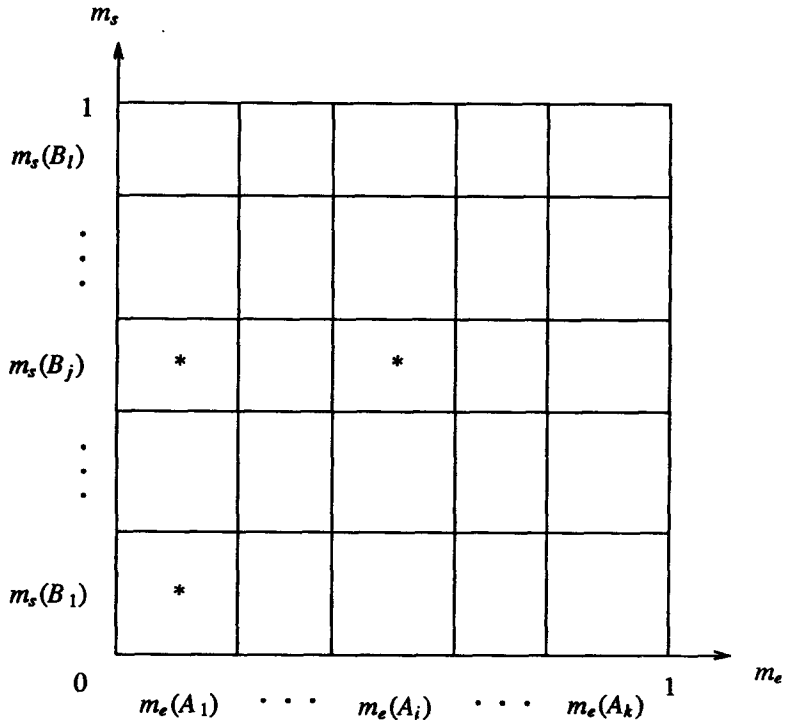


Figure 1. Belief function Bel(o).

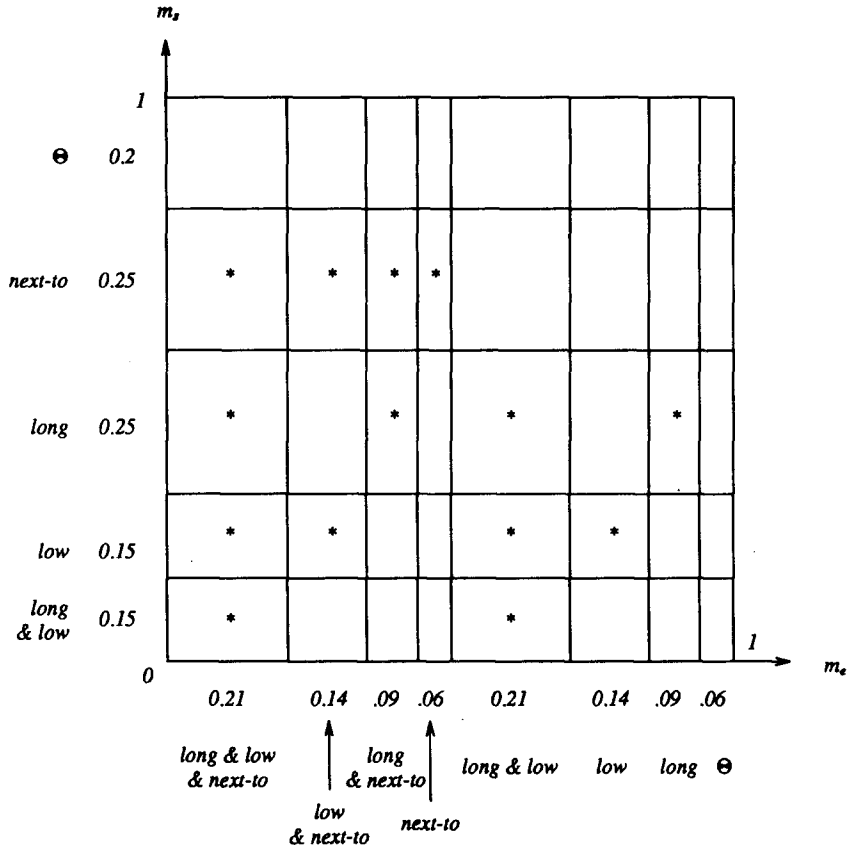


Figure 2. Hypothesis verification.

When the evidential reasoning is applied hierarchically, Bel(o) can be used, for example, together with other pieces of evidence for the higher-level object it implies.

Figure 2 illustrates this type of hypothesis verification. The knowledge of window shutters [m_s(*shutter*)] described earlier is used. Some accumulated evidence (m_e) is assumed. The bpa for the new generated belief function of this hypothesis verification is

$$m(shutter) = 0.21 \times (0.15 + 0.15 + 0.25 + 0.25) + 0.14 \times (0.15 + 0.25) + 0.09 \times (0.25 + 0.25) + 0.06 \times 0.25 + 0.21 \times (0.15 + 0.15 + 0.25) + 0.14 \times 0.15 + 0.09 \times 0.25 = 0.44$$

$$m(\Theta) = 1 - 0.44 = 0.56$$

EXPERIMENTAL RESULTS

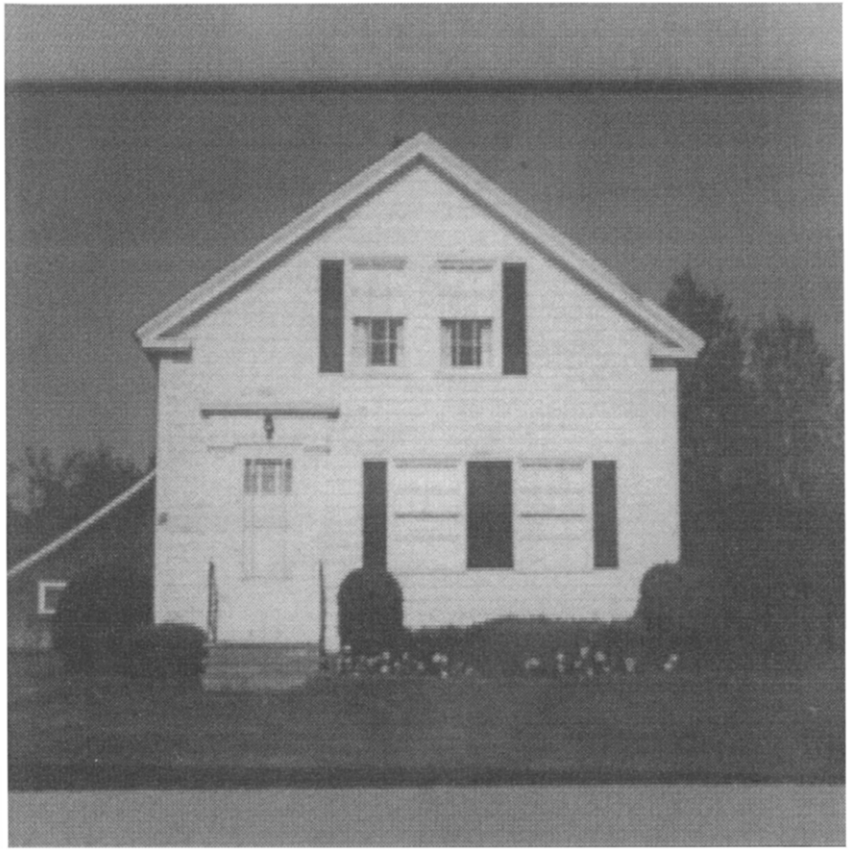
In this section, two test images (Figs. 3*a* and 3*b*) will be used to illustrate the uncertainty management mechanism in our pyramid vision system.

Pyramidal Vision Using Salient Features and Evidential Reasoning

Our vision system runs on a simulated pyramid machine (Table 1). The basic pyramid structure has n levels, each level k ($0 \leq k < n$) has $2^k \times 2^k$ nodes (each a simple processor with its own memory). Each node at level k is hard-wired to its 13 neighbors, that is, one parent, eight siblings, and four children.

(a)

Figure 3. Two test images. (*a*) Building1; (*b*) House.sri.



(b)

Figure 3. *Continued.*

The connectivity between layers of nodes in the pyramid structure makes the hierarchical data flow (bottom-up and/or top-down) very efficient.

The pyramid vision system uses salient features and evidential reasoning [2]. Initially, primitive local features are assessed in parallel by low-level processors and are then used to compute more global and abstract features at consecutively higher levels (bottom-up) in the pyramid. Once certain salient (important) features are extracted, a hypothesis will be generated. Typically the program will then move, in a top-down manner, to lower levels in order to search for more evidence that might verify or deny the hypothesis. Sometimes "lateral search" processes are invoked. The evidential reasoning processes are thus executed in a parallel and hierarchical manner within different nodes and different levels in the pyramid.

Table 1. Configuration of the Simulated Pyramid Machine

Level	Size	Local Memory (Bits)
0	1 × 1	2048
1	2 × 2	2048
2	4 × 4	2048
3	8 × 8	2048
4	16 × 16	2048
5	32 × 32	2048
6	64 × 64	1024
7	128 × 128	256
8	256 × 256	64
9	512 × 512	16

Results from the Image "Building1"

This section illustrates a D-S evidential reasoning example with the image "Building1" (Fig. 3*a*). Only the window recognition part will be described. The original digitized image has a resolution of 512 × 512 pixels. The pyramidal multilevel median filtering method is applied as suggested in Ref 1. As a result, a filtered image with a reduced resolution 128 × 128 is obtained. It is stored at level 7 of the pyramid.

The algorithms for analyzing building images can be found in Ref. 2. Briefly, micro-edges with eight possible directions are first detected with SIMD operations at level 7 nodes. Next, short edges are extracted by level 6 nodes and long edges by level 5 nodes with parallel and hierarchical pyramid operations. Pairs of parallel long edges with opposite directions (e.g., long edges with 0° and 180°, or 90° and 270°) have been found to be good clues for windows with rectangular shapes. Thus such pairs of horizontal long edges are used to predict the possible window areas.

Figure 4 shows the locations of all the possible window areas predicted in the office building image. [Forty-six possible window areas are hypothesized by the program. For ease of demonstration, the real window areas are labeled W1–W12, and other (false) areas are labeled by numbers 1–34. Since the procedure looks for only horizontal long edges, six narrow windows in Fig. 3*a* are not located in this step.]

Apparently, additional evidence is needed and should be accumulated to distinguish the true window areas from others. The following are illustrations of how multiple features (evidence) can be employed to help the process of evidential reasoning.

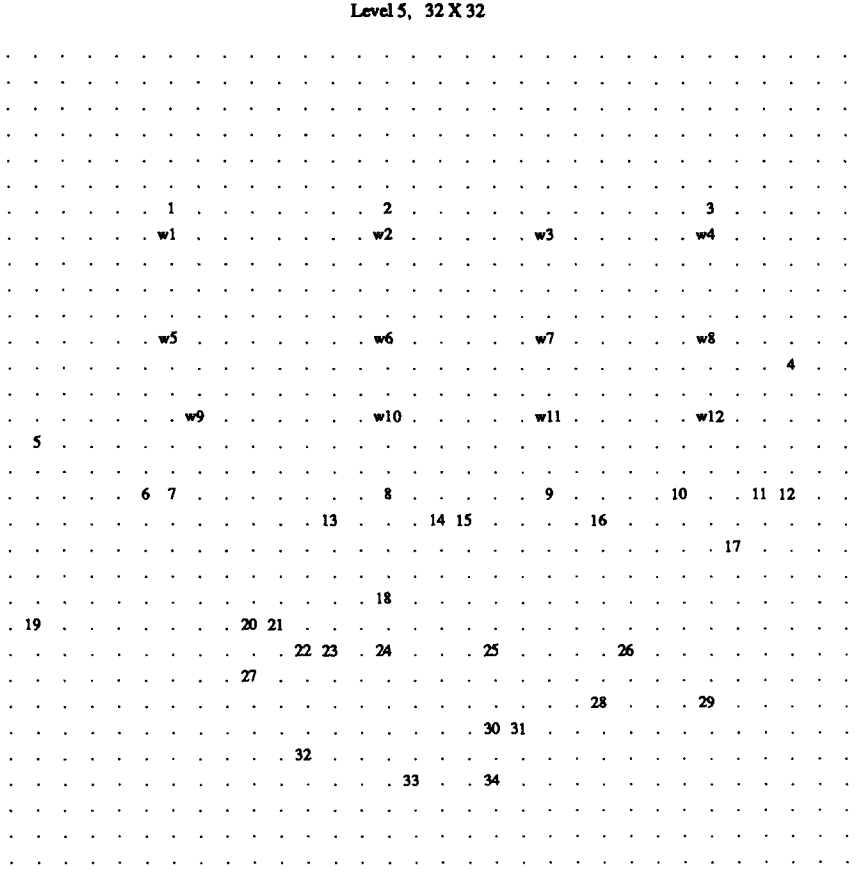


Figure 4. Possible window areas in "Building1."

EVIDENCE OF SHAPES AND TEXTURES Several features of the possible window areas are assessed at level 5. Belief values (Bel) for the window features are assigned.

1. The elongation of the area. The *elongation* of the rectangular areas is defined as *length/width* if *length* $\geq$ *width*; *width/length*, otherwise. Under this definition, the elongation values of typical windows are likely to be small positive numbers larger than 1. The following belief values are given for the elongation of windows:

$$\text{Bel}(elong) = \begin{cases} 0.5 & \text{if } elongation \leq 3 \\ 0.3 & \text{if } 3 < elongation \leq 5 \\ 0 & \text{otherwise} \end{cases}$$

2. The texture of the area. The interior texture of window areas can often be measured by the number of micro-edges. For opened windows or big glass windows, there are very few micro-edges inside the areas. In the case of windows with a few panes or half-closed window shades, there will be overwhelmingly more horizontal and vertical edges (micro-edge 0, 2, 4, 6) than diagonal edges (micro-edge 1, 3, 5, 7). Hence the belief value for the window texture measure is

$$\text{Bel}(\textit{text}) = \begin{cases} 0.4 & \text{if } \textit{edgedness} < 0.1 \\ 0.4 & \text{if } hv/d < 4 \\ 0.2 & \text{if } 2 \le hv/d < 4 \\ 0 & \text{otherwise} \end{cases}$$

where *edgeness* = number-of-edges/area-of-region, and hv/d is the ratio of the total number of horizontal and vertical edges to the number of diagonal edges. The micro-edges are extracted at level 7. The statistical measures are gathered in one bottom-up pass in the pyramid [2].

3. The vertical edges that serve as the left and right boundaries of the windows. Perfect boundaries cannot always be extracted. Bel(*left-bound*) and Bel(*right-bound*) are thus used in a similar way to assess the evidence of the existence of boundaries.

The above set of belief functions are given certain discrete values based on some a prior knowledge. The magnitudes of these values are not crucial to the window recognition. We also experimented with some other sets of belief values. The final recognition result was not sensitive to those initial selections.

The belief functions initially assigned to the single features are all "simple belief functions." While Bel is assigned to the proposition (e.g., *along*) supporting the window hypothesis, the value (1 − Bel) is assigned to Θ to represent the uncertainty.

After all four features have been assessed individually, these simple belief functions are combined to derive a new belief function:

$$\text{Bel}(\textit{elong}) \oplus \text{Bel}(\textit{text}) \oplus \text{Bel}(\textit{left-bound}) \oplus \text{Bel}(\textit{right-bound})$$

whose mass distribution was denoted by m_e in the previous section. For verifying window hypotheses, the combined evidence is compared to the knowledge source, having the following probability mass distribution m_{s_1}:

$$m(\textit{elong}) = 0.15, \quad m(\textit{text}) = 0.20, \quad m(\textit{bound}) = 0.35, \quad m(\Theta) = 0.3$$

The existence of boundaries is thought to be the most convincing evidence; thus a big portion of the total mass is assigned to the proposition *bound*. A certain amount of the mass is attributed to Θ to reflect the incompleteness of this knowledge source.

The results of this hypothesis verification are the belief function Bel(*wnd*) for

the hypothesized window areas. The belief values for single features and the resulting Bel(*wnd*) are shown in Table 2. The 18 possible window areas in Table 2 are the ones that obtain higher Bel(*wnd*) values. The remaining 28 (less possible) areas are not shown.

EVIDENCE ABOUT THE GEOMETRICAL RELATIONS OF WINDOWS At this step, the geometrical relations between windows are used to further improve the results. Since windows of a building are usually arranged in a horizontal or vertical alignment, all the possible window areas can search for possible sibling window areas vertically and horizontally. If a vertical sibling is found, the Bel(*v-sibl*) will be set of 0.6. In the same way, Bel(*h-sibl*) will be set. The combined belief function Bel(*wnd*) $\oplus$ Bel(*h-sibl*) is checked with the knowledge source for verifying the hypothesis that the window area in question is one of the windows in a building. The knowledge source is represented by m_{s_2}:

$$m(wnd) = 0.4, \qquad m(v\text{-}sibl) = 0.2, \qquad m(h\text{-}sibl) = 0.2,$$

$$m(v\text{-}sibl \ \& \ h\text{-}sibl) = 0.2$$

A part of the mass values is assigned to the conjunction of *v-sibl* and *h-sibl* to emphasize the change of the coexistence of both types of siblings. Because this knowledge is though to be very certain, no mass is assigned to Θ.

The resulting belief values are Bel$'$(*wnd*) in Table 2. The previous results have been enhanced by the use of the geometrical relations between expected windows. While the belief values for non-window areas decrease (from 0.335 to 0.134 for areas 4 and 15), the values for all the real window areas get increased. Thus the contrast between the window and non-window areas becomes better.

HANDLING OF CONFLICTING EVIDENCE One of the advantages of the Dempster–Shafer set-theoretical theory of evidence is the ability to accommodate conflicting evidence. The past part of our example will illustrate this capability by applying the reasoning mechanism to independent observations of conflicting evidence.

The simplicity of the image "Building1," for example, the flat roof and the clear outline, makes it very easy for the program to find the building's boundaries. The prediction can even be made using fairly low resolution images at level 5(32 $\times$ 32) in the pyramid. Any areas falling outside the boundaries could hardly be windows of the building. Therefore a proposition *non-wnd* is introduced as the negation of the proposition *wnd*. Areas outside of the building boundaries are assigned Bel(*non-wnd*) = 0.5, and other areas are assigned Bel(*non-wnd*) = 0. The remainder of the unit mass is assigned to Θ. The new belief about *non-wnd* is combined with the belief about *wnd*, *v-sibl*, and *h-sibl* to derive

$$\text{Bel}(non\text{-}wnd) \oplus \text{Bel}(wnd) \oplus \text{Bel}(v\text{-}sibl) \oplus \text{Bel}(h\text{-}sibl)$$

Table 2. Belief Values for Possible Windows

	Possible Window Areas												
	W1–6	W7	W8	W9	W10	W11	W12	4	5	9	15	17	18
Bel(*elong*)	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.3	0.5	0.5	0.3	0.5	0.5
Bel(*text*)	0.4	0.2	0.4	0.4	0.4	0.4	0.4	0.4	0	0.4	0.4	0	0
Bel(*lt-bound*)	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0	0.1	0	0	0.6	0.3
Bel(*rt-bound*)	0.6	0.6	0.6	0.3	0.1	0.6	0.6	0.6	0.3	0.3	0.6	0	0.1
Bel(*wnd*)	.449	.409	.449	.407	.379	.449	.449	.335	.205	.262	.335	.285	.205
Bel(*v-sibl*)	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0	0	0.6	0	0.6	0.6
Bel(*h-sibl*)	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0	0.6	0	0	0	0
Bel'(*wnd*)	.492	.475	.492	.475	.462	.492	.492	.134	.203	.225	.134	.234	.203
Bel(*non-wnd*)	0	0	0	0	0	0	0	0.5	0.5	0	0	0	0.5
Bel″(*wnd*)	.492	.475	.492	.475	.462	.492	.492	.080	.166	.225	.134	.234	.166

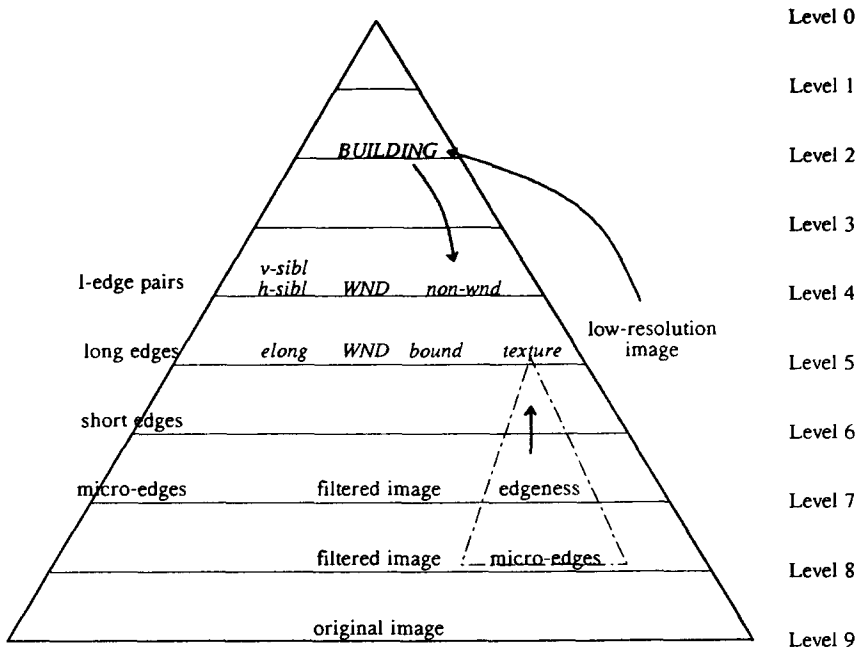


Figure 5. Multilevel processes in the analysis of "Building1."

Finally, the same knowledge source with m_{s_2} is utilized to verify the *wnd* hypothesis. As expected, the belief values for those outside nodes (4, 5, and 18) are further decreased [see Bel″(*wnd*) in Table 2].

The final values Bel″(*wnd*) show that the program is very confident about the existence of the 12 real windows. Because the program found strong evidence as to the number and the arrangement of the windows at this level, it combined this with other evidence (e.g., the shape and outline of the building) and succeeded in recognizing the object in the building scene as an office or apartment building.

Figure 5 concludes this example by depicting the features extracted in different levels in the pyramid. The arrow pointed to *Building* indicates that a high-level hypothesis (the outline of the building) is generated from low-resolution images (32 × 32) at level 5. The arrow pointing to *non-wnd* indicates that the conflicting evidence (for *non-wnd*) comes from the high-level process. The arrow pointing to *texture* indicates that the data for the texture measure (*edgeness*) flow in a bottom-up away from level 8 to level 5. The rest of the data flow is not indicated by arrows.

Results from the Image "House.sri"

This section briefly illustrates how our D-S method analyzes the "window-assemblies" (windows and shutters) in the image "House.sri" also having an original resolution 512 × 512 (Fig. 3*b*).

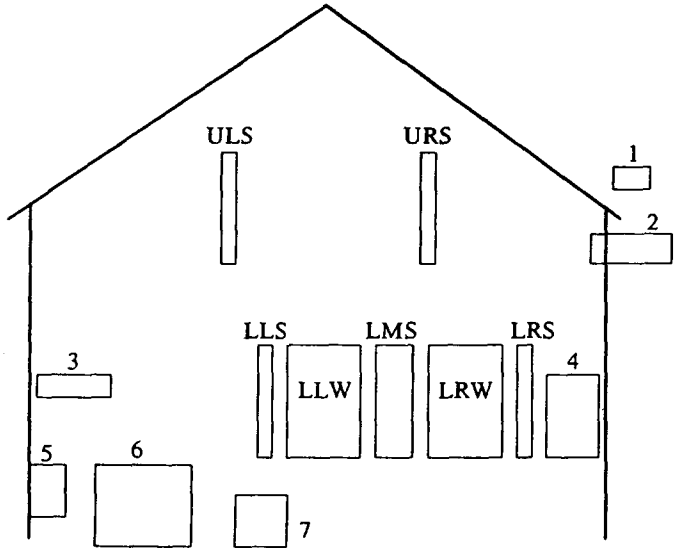


Figure 6. Possible window or shutter areas.

Micro-edges are again extracted at level 7. Using the same set of low-level processes and pairs of vertical long edges as salient features for windows and shutters, 14 possible window or shutter areas are located. Figure 6 is a graphical display of these 14 areas. For ease of illustration, real window and shutter areas in "House.sri" are labeled LLW, LLS, etc. Others are labeled from 1 to 7. (The program missed the top windows at this point and recognized them at a later stage.)

Similar to what we did in the analysis of "Building1," features of elongation, texture, and boundary of the areas are examined.

Note that the evidence on the proposition "window" will also be used as the evidence on the proposition "non-shutter" (conflicting to "shutter"), and vice versa.* For examples, if *elongation* = 4, then

$$\text{Bel}(elong(sht)) = \text{Bel}(elong(non\text{-}wnd)) = 0.5$$

and

$$\text{Bel}(elong(wnd)) = \text{Bel}(elong(non\text{-}sht)) = 0$$

* In this case, the frame of the discernment Θ consists of "window," "shutter," and "other-objects." Since no mass value is assigned to "other-objects," Bel(*elong non-sht*) = Bel(*elong*(*wnd*)), and vice versa.

The simple belief functions [Bel(*elong*(*sht*)), Bel(*text*), etc.] are combined to derive a new belief function. For *shutter* it is

$$\text{Bel}(elong(sht)) \oplus \text{Bel}(elong(non\text{-}sht)) \oplus \text{Bel}(text) \oplus \text{Bel}(top\text{-}bound) \oplus \text{Bel}(bottom\text{-}bound)$$

whose mass distribution is denoted by $m_e(sht)$. In the same way, $m_e(wnd)$ is derived.

The mass distributions for the knowledge sources used to verify the shutter or window hypotheses are:

$m_{s_1}(sht)$: $m(elong(sht)) = 0.4$ $m(text) = 0.15$
$m(bound) = 0.25$ $m(\Theta) = 0.2$

$m_{s_1}(wnd)$: $m(elong(wnd)) = 0.4$ $m(text) = 0.15$
$m(bound) = 0.25$ $m(\Theta) = 0.2$

Since the "elongation" of the area is an important feature used to discriminate shutters from windows in this example, it is emphasized by assigned a large portion of the total mass to the proposition *elong*. The resulting belief functions are Bel(*sht*) and Bel(*wnd*) as shown in Table 3.

In house scenes, windows and shutters in "window-assemblies" are usually horizontal siblings. The belief function Bel(*h-sibl*) is introduced to represent this geometrical relation. When this is incorporated into the reasoning process, the final belief functions are Bel′(*sht*) and Bel′(*wnd*).

COMPARISONS OF REASONING MECHANISMS

The previous sections described the D-S evidential reasoning mechanism and its application in real-world image analysis. An immediate question is: "Does it work better than other reasoning mechanisms?" In this section empirical comparisons are made among reasoning mechanisms using the D-S theory, the Bayesian formalism, and a simple weight combination method for the same set of image data. Let us first concentrate on the image "House.sri."

Methods for Decision Making

In this example the vision system is expected to identify the most probable window and shutter areas. At least two simple methods may be recommended for decision making.

Method 1: Choose the label with the highest Bel value.

For each possible window or shutter area, examine its Bel′(*sht*), *Bel′*(*wnd*),

Table 3. Belief Values for Features in "House.sri"

	Possible Window or Shutter Areas													
	ULS	URS	LLS	LMS	LRS	LLW	LRW	1	2	3	4	5	6	7
Bel(*elong*(*sht*))	0.5	0.5	0.5	0.3	0.5	0	0	0	0.3	0.3	0	0	0	0
Bel(*elong*(*wnd*))	0	0	0	0.3	0	0.5	0.5	0.5	0.3	0.3	0.5	0.5	0.5	0.5
Bel(*text*)	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0	0	0	0.4	0.4	0.4	0.4
Bel(*top-bound*)	0.3	0.6	0.6	0.6	0.6	0.6	0.6	0.1	0.3	0	0	0	0.3	0.1
Bel(*btm-bound*)	0.3	0.6	0	0.6	0.6	0.3	0.1	0.1	0.1	0	0	0	0.1	0.3
Bel(*sht*)	.388	.470	.410	.357	.470	.240	.220	.048	.183	.092	.060	.060	.154	.154
Bel(*wnd*)	.188	.270	.210	.357	.270	.440	.420	.248	.183	.092	.260	.260	.354	.354
Bel(*h-sibl*)	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0	0	0	0.6	0	0	0
Bel′(*sht*)	.451	.509	.467	.429	.509	.348	.334	.034	.128	.064	.222	.042	.108	.108
Bel′(*wnd*)	.311	.369	.327	.429	.369	.488	.474	.174	.128	.064	.362	.182	.248	.248

and Bel′(*others*). Although the Bel′(*others*) are not shown in Table 3, a rough assumption could be made for this discussion that Bel′(*others*) is low for a true window or shutter area and high for other areas. it can be seen from the table that all the areas would be labeled correctly except that the LMS will end up in a draw between *sht* and *wnd*.

Method 2: Choose a threshold (θ) for each label.

An area will be labeled as a window or shutter if its associated belief value is higher than the chosen threshold. As shown in Table 3, the real shutter areas obtain the highest Bel′(*sht*) values (0.451, 0.509, 0.467, 0.429, and 0.529); the real window areas have the highest Bel′(*wnd*) values (0.488, 0.474). Therefore, it is not difficult to assign a unique label to each area. In the case where the initial threshold was too low, for example, 0.4 for both *sht* and *wnd*, the area LMS would be labeled both as *sht* and *wnd*. An adjustment would be needed to raise some of the thresholds. For instance, if θ(*wnd*) is raised to about 0.45, then the unique decision would be made.

Reasoning Based on the Bayesian Formalism

For empirical comparison an experiment was run replacing our D-S evidential reasoning mechanism with a method based on the Bayesian formalism.

THE REASONING MECHANISM The same set of features (*elongation*, *texture*, etc.) is employed. The reasoning process will compute the posterior probability for the hypothesized object (H) given multiple uncertain evidence (E_i). The mechanism for updating Bayesian probabilities suggested by Duda et al. [12] is adopted.

The *prior odds* on a hypothesis H are defined to be

$$O(H)=\frac{P(H)}{P(\bar{H})}=\frac{P(H)}{1-P(H)} \tag{1}$$

and the *posterior odds* to be

$$O(H|E)=\frac{P(H|E)}{P(\bar{H}|E)}=\lambda O(H) \tag{2}$$

when the evidence E is known to be true, and the *likelihood ratio* λ is defined as

$$\lambda=\frac{P(E|H)}{P(E|\bar{H})} \tag{3}$$

In a strictly analogous fashion, if E is known to be false, then

$$O(H|\bar{E})=\frac{P(H|\bar{E})}{P(\bar{H}|\bar{E})}=\bar{\lambda}O(H) \tag{4}$$

where $\bar{\lambda}$ is defined as

$$\bar{\lambda}=\frac{P(\bar{E}|H)}{P(\bar{E}|\bar{H})} \tag{5}$$

If an uncertain evidence E_i' is observed, then $P(E_i|E_i')$ is used to represent the probability that E_i is true under the observation of E_i'. It was assumed in Ref. 12 that

$$P(H|E_i')=P(H|E_i)P(E_i|E_i')+P(H|\bar{E}_i)P(\bar{E}_i|E_i') \tag{6}$$

For combining multiple uncertain evidence, an effective likelihood ratio λ_i' is defined for each single feature by

$$\lambda_i'=\frac{O(H|E_i')}{O(H)} \tag{7}$$

Assuming that the E_i are conditionally independent, the posterior odds given E_1', $\cdots$, E_n' are

$$O(H|E_1', \cdots, E_n')=\left[\prod_{i=1}^{n}\lambda_i'\right] O(H) \tag{8}$$

The reasoning process works as follows. The prior odds $O(H)$ for possible objects (*window, shutter,* or *others* in this example) should be given. For each observed (uncertain) feature E_i' (e.g., *elongation, texture*), the λ_i*YPP* and $\bar{\lambda}_i$ are also given. The system will need some subjective knowledge for these λ_i and $\bar{\lambda}_i$. Alternatively, they may be generated from previous statistical data. The $O(H|E_i)$ and $O(H|\bar{E}_i)$ can be derived from Eqs. (2) and (4). The $P(H|E_i)$ and $P(H|\bar{E}_i)$ can consequently be obtained. [$P = O/(O + 1)$ and $O = P/(1 - P)$, the conversions between P and O, are used here and hereafter.] For each observed feature the initial belief value Bel_i in the D-S method is now simply used as $P(E|E_i')$, which will enable simple comparison between these two methods. It follows that $P(\bar{E}|E_i') = 1 - P(E|E_i')$. With Eq. (6) the $P(H|E_i')$ is calculated and will then be used to derive $O(H|E_i')$. Subsequently, λ_i' for feature i is calculated from (7). After obtaining all λ_i, the posterior odds $O(H|E_1', \cdots, E_n')$ can be computed from (8). Finally, the $P(H|E_1', \cdots, E_n')$ for each possible object can be derived.

Example: The impact of the observed *texture* on the hypothesized object *window*. Let us consider the small set of possible objects {*wnd, sht, others*}. Suppose the prior probabilities $P(wnd) = 0.167$, $P(sht) = 0.33$, and $P(others) = 0.5$. For the feature *texture,* $\lambda_{text} = 2$ and $\bar{\lambda}_{text} = 0.8$. Assume that the observation on texture has 40% certainty that there is some window texture, that is, $P(text|text') = 0.4$.

$$O(wnd)=\frac{0.167}{1-0.167}=0.2$$

$$O(wnd|text) = 2 \times 0.2 = 0.4 \qquad P(wnd|text) = \frac{0.4}{1+0.4} = 0.286$$

$$O(wnd|\overline{text}) = 0.8 \times 0.2 = 0.16 \qquad P(wnd|\overline{text}) = \frac{0.16}{1+0.16} = 0.138$$

$$P(wnd|text') = 0.286 \times 0.4 + 0.138 \times (1-0.4) = 0.197$$

$$O(wnd|text') = \frac{0.197}{1-0.197} = 0.245$$

$$\lambda'_{text} = \frac{0.245}{0.2} = 1.227$$

Similar computations are made to obtain λ'_i for other observed features (*elong, top-bound, btm-bound,* and *h-sibl*). Hence, the final updated probabilities can be derived from Eq. (8). Table 4 shows the results of reasoning with this implementation of Bayesian formalism. The first six rows are $P(E|E'_i)$ for observed features. The posterior probabilities after the combination of the multiple evidence are denoted by $P'(wnd)$ and $P'(sht)$, respectively.

For this computation the prior probabilities are chosen to be $P(wnd) = 0.167$, $P(sht) = 0.333$, and $P(others) = 0.5$. The reasons for this are: (1) For the house pictures we are analyzing, each window has two shutters. Thus, prior $P(sht) = 2 \times P(wnd)$. (2) Usually, after initial image analysis steps the number of hypothesized possible window or shutter areas is much larger than the number of actual window or shutter areas. Therefore, $P(others)$ is assigned a larger prior probability.

ANALYSIS OF THE RESULTS The results are comparable with those of D-S method. All five shutter areas have higher $P'(sht)$ than their $P'(wnd)$ (e.g., $0.552 > 0.211$ for ULS), and two window areas have higher $P'(wnd)$ than their $P'(sht)$ (e.g., $0.487 > 0.430$ for LLW). The correct decision would be made if decision method 1 were the choice. However, there would be some problem if decision method 2 were used. Note that area LMS has an unexpectedly high $P'(wnd)$ (0.495), which is even higher than the $P'(wnd)$ values of the two true window areas. It is not apparent how the system can avoid identifying LMS as a *wnd* while using decision method 2.

As expected, the values of the λ_i and $\bar{\lambda}_i$ will have some impact on the posterior probabilities. Also, the belief values in the D-S method might be too low to be directly adopted as reasonable probability values. We experimented with many different groups of these values. Although the magnitudes of the posterior probabilities change substantially, the relative measure of these probabilities is not significantly affected. Namely, the areas having higher P values always have

Table 4. Probability Values for Features in "House.sri"

	Possible Window or Shutter Areas													
	ULS	URS	LLS	LMS	LRS	LLW	LRW	1	2	3	4	5	6	7
P(elong(sht))	0.5	0.5	0.5	0.3	0.5	0	0	0	0.3	0.3	0	0	0	0
P(elong(wnd))	0	0	0	0.3	0	0.5	0.5	0.5	0.3	0.3	0.5	0.5	0.5	0.5
P(text)	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0	0	0	0.4	0.4	0.4	0.4
P(top-bound)	0.3	0.6	0.6	0.6	0.6	0.6	0.6	0.1	0.3	0	0	0	0.3	0.1
P(btm-bound)	0.3	0.6	0	0.6	0.6	0.3	0.1	0.1	0.1	0	0	0	0.1	0.3
P(h-sibl)	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0	0	0	0.6	0	0	0
P'(sht)	.552	.708	.541	.641	.708	.430	.371	.084	.166	.107	.201	.094	.148	.148
P'(wnd)	.211	.345	.195	.495	.345	.487	.419	.090	.087	.050	.221	.099	.166	.166

higher P' values and vice versa. Noticeably, the $P'(wnd)$ for LMS is always higher than the $P'(wnd)$ for LRW.

A Simple Weight Combination Method

An experiment was also run by using a simple weight combination method that is used by some simple perception systems. The initial belief values used in the D-S method are now simply treated as weights (WT). Weights for the features that support the proposition, for example, *shutter*, are summed up. Weights for the features that support the negation of the propositions are subtracted from the sum. Thus, for shutters,

$$\mathrm{WT}(sht) = \mathrm{WT}(elong(sht)) - \mathrm{WT}(elong(\text{non-sht})) + \mathrm{WT}(text) + \mathrm{WT}(top\text{-}bound) + \mathrm{WT}(bottom\text{-}bound)$$

With the consideration of the possible support from siblings,

$$WT'(sht) = WT(sht) + WT(h\text{-}sibl)$$

In the same way, WT(wnd) and WT$'$(wnd) are obtained.

Table 5 lists all these weights for "House.sri."

The result is reasonably good. All the real shutter areas get higher WT(sht) (namely, 2.1, 2.7, 2.1, 2.2, and 2.7) than non-shutter areas. Also, two lower window areas get high WT$'$(wnd) values, 2.4 and 2.2. However, there is a problem similar to one in the Bayesian method: WT$'$(wnd) for the shutter LMS is 2.2, which is as high as the WT$'$(wnd) for one of the windows LRW.

A Graphical Comparison

Figure 7 consists of histograms that depict the final belief on the hypothesis *window* derived with three different reasoning mechanisms. Data from Tables 3–5 are used. As explained before, 14 areas are candidate window or shutter areas in the image "House.sri." For ease of viewing, the real shutters and windows are represented by boxes in different shades. The horizontal axes indicate the belief values, probabilities, or weights. Since no normalization is attempted, the magnitudes of Bel$'$, P', and WT$'$ are not simply compared. The histograms take the minimum and maximum values (i.e., 0.034 and 0.509 for belief values, 0.095 and 0.495 for probabilities, -0.3 and 2.7 for weights) and display them with an approximately equal width, so that the discrimination power of the different reasoning methods can be shown by the horizontal distances—for example, how far windows are separated from non-windows. A good reasoning process should be able to obtain the highest Bel$'$(wnd) [or $P'(wnd)$, WT$'$(wnd)] for the two true window areas. In Figure 7*a* the two true window areas LLW and LRW have the highest Bel$'$(wnd) values and thus are separated from the "false" window areas. Figures 7*b* and *c* show that one shutter area is not separated from the true window areas.

Table 5. Weights for Features in "House.sri"

	Possible Window or Shutter Areas													
	ULS	URS	LLS	LMS	LRS	LLW	LRW	1	2	3	4	5	6	7
WT(*elong*(*sht*))	0.5	0.5	0.5	0.3	0.5	0	0	0	0.3	0.3	0	0	0	0
WT(*elong*(*wnd*))	0	0	0	0.3	0	0.5	0.5	0.5	0.3	0.3	0.5	0.5	0.5	0.5
WT(*text*)	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0	0	0	0.4	0.4	0.4	0.4
WT(*top-bound*)	0.3	0.6	0.6	0.6	0.6	0.6	0.6	0.1	0.3	0	0	0	0.3	0.1
WT(*btm-bound*)	0.3	0.6	0	0.6	0.6	0.3	0.1	0.1	0.1	0	0	0	0.1	0.3
WT(*sht*)	1.5	2.1	1.5	1.6	2.1	0.8	0.6	−0.3	0.4	0	−0.1	−0.1	0.3	0.3
WT(*wnd*)	0.5	1.1	0.5	1.6	1.1	1.8	1.6	0.7	0.4	0	0.9	0.9	1.3	1.3
WT(*h-sibl*)	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0	0	0	0.6	0	0	0
WT′(*sht*)	2.1	2.7	2.1	2.2	2.7	1.4	1.2	−0.3	0.4	0	0.5	−0.1	0.3	0.3
WT′(*wnd*)	1.1	1.7	1.1	2.2	1.7	2.4	2.2	0.7	0.4	0	1.5	0.9	1.3	1.3

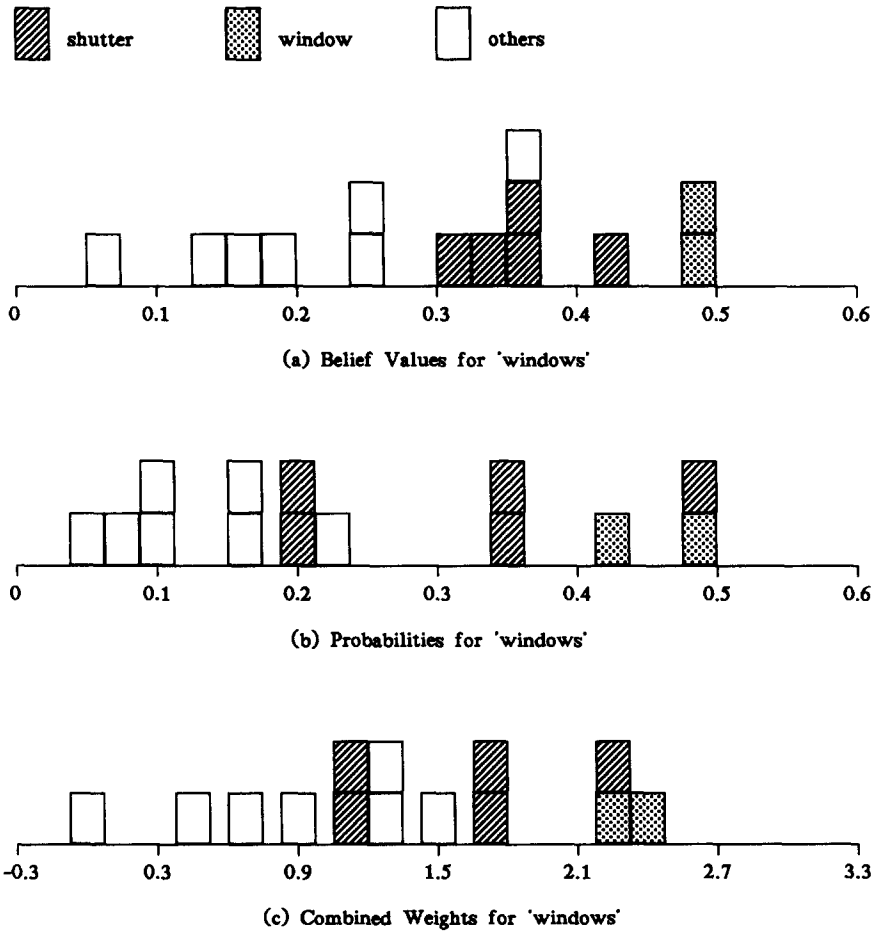


Figure 7. A graphical comparison.

The three mechanisms in comparison work in rather different ways. To make a good empirical comparison, many more studies and tests are needed. For example, the impact (sensitivity) of the selection of initial values (Bel, P, λ, etc.) on the final results should be tested more carefully. We believe that the following are the factors that do make our evidential reasoning approach perform better in the "House.sri" example:

1. The D-S theory offers a natural way to combine evidence. Our extended use of the theory facilitates the management of both uncertain data and uncertain knowledge in the evidence accumulation and hypothesis verification steps.
2. The D-S evidential reasoning approach incorporates the knowledge of salient features and has a nice way of emphasizing the importance of these

features for object recognition. For example, the elongation is emphasized by assigning a large amount of mass to it in m_3.

3. The evidence of *elong(wnd)* is also used as *elong(non-sht)*. This is sound and only feasible in the D-S and weight combination approaches. It helped in discriminating LMS from the two windows.

Similar comparisons were also made in the analysis of the image "Building1." All three methods (the D-S mechanism, the Bayesian formalism, and the simple weight combination) have no difficulty in separating 12 real windows from 34 non-windows. The results are fairly comparable.

CONCLUSION

This paper describes how the Dempster–Shafer theory of evidence is used in a massively parallel hierarchical pyramid system for computer vision. The vision program was originally designed to recognize complex real-world objects (e.g., houses, office buildings, neurons) using a structure of micromodular production rule–like transforms that are applied in a combined bottom-up and top-down flow. Evidential reasoning is embedded at several stages in this program's processes. The knowledge representation scheme described in this paper uses uncertain and incomplete world knowledge. It is modular, flexible, and easy to update. The hypothesis-verification mechanism extends the applications of the belief functions and Dempster's combination rule in a relatively efficient and powerful manner. As shown by our examples, this evidential reasoning mechanism serves to disambiguate and enhance the program's judgments about objects. Preliminary tests indicate that it is very comparable with other methods based on the Bayesian formalism and the simple weight combination method. In the "House.sri" examples it does improve the program's performance.

References

1. Li, Z. N., and Uhr, L., A pyramidal approach for the recognition of neurons using key features, *Pattern Recog.* **19**(1), 55–62, 1986.

2. Li, Z. N., and Uhr, L., Pyramid vision using key features to integrate image-driven bottom-up and model-driven top-down processes, *IEEE Trans. Syst. Man, Cybern.* **SMC-17**(2), 250–263, 1987.

3. Garvey, T. D., Lowrance, J. D., and Fischler, M. A., An inference technique for integrating knowledge from disparate sources, *Proc. 7th IJCAI 319–325, 1981.*

4. Lowrance, J. D., Dependency-graph models of evidential support, PhD Thesis, University of Massachusetts at Amherst, 1982.

5. Gordon, J., and Shortliffe, E. H., The Dempster–Shafer theory of evidence, in

Rule-Based Expert Systems (B. G. Buchanan and E. H. Shortliffe, Eds.), Addison-Wesley, Reading, Mass., 1984.

6. Lu, S. Y., and Stephanou, H. E., A set-theoretic framework for the processing of uncertain knowledge, *Proc. 4th National Conf. on AI*, 216–221, 1984.
7. Wesley, L. P., Lowrance, J. D., and Garvey, T. D., Reasoning about control: an evidential approach, SRI Tech. Note 324, 1984.
8. Wesley, L. P., Evidential knowledge-based computer vision, *Opt. Eng.* **25**(3), 363–379, 1986.
9. Zhang, M. C. and Chen, S. S., Evidential reasoning in image understanding, *Proc. AAAI 87 Workshop on Uncertainty in AI,* 340–346, 1987.
10. Dempster, A. P., Upper and lower probabilities induced by a multivalued mapping, *Ann. Math. Stat.* **38**, 325–339, 1967.
11. Shafer, G., *A Mathematical Theory of Evidence,* Princeton University Press, Princeton, N. J., 1976.
12. Duda, R. O., Hart, P. E., and Nilsson, N. J., Subjective Bayesian methods for rule-based inference systems, *Proc. AFIPS 1976 National Computer Conf.* **45**, 1075–1082, 1976.